

Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 15:19:33

PAGE 1

REFERENCE NO: 259

This contribution was submitted to the National Science Foundation as part of the NSF CI 2030 planning activity through an NSF Request for Information, https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf17031. Consideration of this contribution in NSF's planning process and any NSF-provided public accessibility of this document does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government. The content of this submission is protected by the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>).

Author Names & Affiliations

- Lisa Arafune - Coalition for Academic Scientific Computation (CASC)

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Academic Scientific Computation/HPC and Research Computing

Title of Submission

Coalition for Academic Scientific Computation (CASC) response to NSF RFI

Abstract (maximum ~200 words).

The four topics presented here reflect general consensus on top priorities among many CASC member institutions concerning challenges faced by the academic HPC and research computing communities that the NSF should consider when creating its own strategies and planning for the next decades.

These topics, in no particular order, reflect the cyberinfrastructure challenges and needs that CASC believes to be critical for ensuring scientific and engineering advances over the next decade and beyond.

? The pyramid model of computational resources (scaling from individual PIs and campuses at the base to leadership class facilities at the top) should continue to play a major part in shaping the support of researcher computational and data needs.

? Developing and ensuring a stable and rewarding career path for the professional workforce that is involved in all facets of cyberinfrastructure is critical.

? Shaping the ecosystem for storing, preserving, curating, finding and managing science and engineering data is also essential.

? Software is a vital part of cyberinfrastructure. Elevating the practices, procedures and various forms of training for developing, distributing, and maintaining scientific software to ensure usable, scalable, and sustainable algorithms and applications is critical to the community.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

The pyramid model of computational resources (scaling from individual PIs and campuses at the base to leadership class facilities at the top) should continue to play a major part in shaping the support of researcher computational and data needs.

It is important to create a balanced and cohesive data and computational ecosystem that reflects the structure of the pyramid (known as the Branscomb pyramid) and recognizes the importance of data and computational science at all levels, from individual PIs and institutions up

Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 15:19:33

PAGE 2

REFERENCE NO: 259

to the regional and national (and international) scale. This should not be the responsibility of NSF alone, but should be undertaken as a collaboration among individual PIs, campuses and all government agencies supporting science and engineering research. To remain competitive, US scientists need access to ever increasing and rapidly changing data and computational cyberinfrastructure (CI), and the NSF should help provide those resources – at the campus, regional, and national levels. Ideally, strategic investments at the national level should encourage further local investment and enable local institutions to integrate into the national CI ecosystem. The exascale initiative is a current driver of much of the innovation happening in cyberinfrastructure, and the NSF should continue to play a leading role in that initiative. Concurrently, the NSF should help provide computational resources for scientific areas or applications that do not require exascale capabilities, as well as continued exploration of additional computing and data paradigms.

Cloud resources – both academic and commercial - are an important component of the ecosystem, and agencies should consider commercial cloud infrastructure as an integral part of the national cyberinfrastructure. To that end, the NSF should reduce the administrative burden for universities - for example, through special agreements with public cloud providers, or through leveraged purchases. The community appreciates continued opportunities to provide input on roadmaps, plans, or any other areas, to assist the NSF. Developing and ensuring a stable and rewarding career path for the professional workforce that is involved in all facets of cyberinfrastructure is critical.

The CI ecosystem depends heavily and increasingly on an emerging professional workforce. These diverse CI professionals are known by many terms: data scientists, computational scientists, research systems and storage administrators and engineers, research software engineers, and cyberinfrastructure and/or research practitioners/facilitators/consultants/trainers. It is vital to foster the continued development of the entirety of this workforce by supporting research leading to innovative educational programs, curricula, and training. This includes support for programs that establish and maintain long-term career paths and a strong professional community. It is essential to the community to provide continuous periods of stable

employment for these professionals, perhaps by incentivizing institutions to include partial funding for them in grants or research contracts. Together, these activities lead to a larger, more diverse, highly skilled effective workforce.

Shaping the ecosystem for storing, preserving, curating, finding, and managing science and engineering data is also essential.

Crucial issues include: fostering high quality data management plans; harmonizing the large number of community endeavors for sharing disparate data; and the alignment of retention and curation practices (including metadata and assessment) across agencies, such as between NSF and NIH, that facilitate long term interdisciplinary research.

Software is a vital part of cyberinfrastructure. Elevating the practices, procedures and various forms of training for developing, distributing, and maintaining scientific software to ensure usable, scalable, and sustainable algorithms and applications is critical to the community.

Software sustainability and maintenance is a key concern among CASC institutions, and is integral to the long term viability of computational and data science approaches to research and discovery. Initiatives such as the research software engineer (RSE) model in the UK may provide examples for the US.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

See response to Question 1.

Consent Statement

- "I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publically available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>)."